Egocentric and Encyclopedic Doxastic States in Delusions of Misidentification

Abstract

A recent debate in the literature on delusions centers on the question of whether delusions are beliefs or not. In this paper, an overlooked distinction between egocentric and encyclopedic doxastic states is introduced and brought to bear on this debate, in particular with regard to delusions of misidentification. The result is that a more accurate characterization of the delusional subject's doxastic point of view is made available. The patient has a genuine egocentric belief ("This man is not my father"), but fails to have the commonly attributed encyclopedic belief ("My father has been replaced by an impostor").

The Capgras delusion is often described as the delusion that "a loved one has been replaced by an impostor".[1] Along with a number of other delusions (e.g. the Frégoli delusion, mirrored self-misidentification), it is classified as a delusion of misidentification, which is to say that the subject fails to correctly identify someone or something that one would expect them to correctly identify. In the case of the Capgras delusion, the subject's perceptual system seems to be working fine: she is not blind, or agnosic, and makes accurate judgments concerning what things and people look like. And yet, in the presence of, say, her father, the Capgras patient will say, "That man is not my father; he looks exactly like my father, but isn't him."

Some theorists have claimed that delusions generally (Currie 2000), and the Capgras delusion in particular (Currie and Jureidini 2001), are not really beliefs. In this paper, I want to re-examine this claim in the light of a distinction between egocentric and encyclopedic doxastic states. Following Bayne and Pacherie (2005), let's call the view that delusions are beliefs, *doxasticism*, and the view that they are not beliefs, *anti-doxasticism*. My aim is not to conclusively criticize either view, but rather to introduce something that seems to have been overlooked in the debate, and which warrants consideration. In the light of this, one may argue that the Capgras patient believes the egocentric proposition, "This man is not my father", but fails to believe the encyclopedic proposition, which is commonly attributed to the Capgras patient, "A certain man, my father, has been replaced by an impostor". The anti-doxasticists are right that the belief commonly attributed to Capgras patients (e.g.

[1] I use "often described" because we shall see that that ascription is not altogether optimal.

in the common description at the very start of this paper) is not actually held by them. However, there is genuine belief in the delusional misidentification, which involves (or is constituted by) egocentric, context-dependent, belief.

What seems to have contributed to this being overlooked is a certain way of thinking about belief, which, in turn, has influenced the way in which the issue concerning whether delusions are beliefs has been framed. In particular, the main players in the debate tend to think of delusion in terms of a discrete, belief-like state, the delusion, which is a belief if it plays a certain functional role, and fails to be a belief if it plays another role. If, instead, we put aside this way of approaching the issue, which carries several implicit commitments (roughly, commitments to the representational theory of mind) and ask ourselves "How does the subject take things to stand in the world?" we (i) get to the heart of the issue and (ii) open up the possibility that some ascriptions may be accurate while others won't be. The challenge then is not primarily to ascertain the functional role of a particular representation, but rather to see if we can intelligibly characterize the patient's cognitive perspective.

In section 1, I present the anti-doxastic position, and why one might be tempted to hold it. In section 2, I outline the distinction between encyclopedic and egocentric doxastic states. In section 3, I consider a general issue about what any subject must do in order to get from an egocentric doxastic state to an encyclopedic one that has the same subject-matter. When that subject-matter is singular, namely, about an individual, *identification* is needed and this is precisely what doesn't happen in the Capgras case. I then consider some implications that this has for our thinking about delusions and belief, and whether delusions are beliefs.

1. Anti-doxasticism

Anti-doxasticism raises a particular instance of a general problem, namely, the problem of distinguishing between the evaluative and constitutive norms of a phenomenon; in other words, distinguishing between doing ϕ badly or well (evaluative) versus doing ϕ or not doing ϕ at all (constitutive). In this instance, ϕ-ing is believing and the anti-doxasticist thinks that the delusional

subject is breaking *constitutive* norms of belief. Unlike the doxasticist, who thinks that delusion is bad believing, the anti-doxasticist thinks that something has gone wrong in such a way that the subject is not believing at all. In spite of what her utterances suggest, the delusional subject doesn't actually believe the content of her delusional claims.

To get a grasp of the distinction between evaluative and constitutive norms, consider an analogy with chess. If, during a game of chess, I move so that my queen can be captured by my opponent with impunity (and as a result she goes on to win the game) that is a *bad* chess move. An evaluative norm (the norm of playing chess well) is broken. If I move my bishop like a rook, however, that's not a chess move at all, let alone a bad one. A constitutive norm (dictated by the rules of chess) is thereby broken. The doxasticist (e.g. Bayne and Pacherie 2005, Bortolotti 2009) will say that delusion is like the former. The anti-doxasticist will say that it's like the latter.

The main proponents of the anti-doxastic position are Gregory Currie and his collaborators (e.g. Currie 2000, Currie and Jureidini 2001). Their view consists in the negative claim that delusions aren't beliefs (anti-doxasticism) and the positive claim about what delusions are instead, namely, they are imaginings that are mistaken by the subject for beliefs (what Bayne and Pacherie (2005) aptly label The "Metacognitive View"). Here I focus exclusively on the negative claim, namely, that delusions aren't beliefs. What they might be instead, I will put aside.

Currie thinks that delusions ought not to be counted as beliefs because although they have the superficial trappings of belief (namely sincere declarative utterances) they lack the right kind of functional role. Since, according to Currie, delusions

(i) are not supported by evidence in their initial formation,

(ii) do not guide action and reasoning, and

(iii) are not open to review in the face of contrary evidence,

they should not be counted as beliefs at all. One way of arguing against this anti-doxasticism is to deny that delusional subjects *actually* infringe these norms. For example, with regard to (i) one can respond that delusions may be based on evidence of a sort, namely on strange experiences (e.g. lack of affective response to familiar faces in the case of Capgras, *locus classicus* Ellis and Young 1990). Maher, for example, hypothesizes that "delusional belief is not being held 'in the face of evidence

strong enough to destroy it', but is being held because evidence is strong enough to support it" (1974, p.99). In response to (ii) one can claim that, although delusions often fail to generate the kinds of actions and emotional responses one might expect, the Capgras delusion (for example) does lead to violence against the impostor in 18% of cases (according to De Pauw and Szulecka 1988), and sometimes of a particularly gruesome sort.[2] (Try telling the man who was stabbed by his previously loving son that the delusion was not acted upon!) Granted, it is much harder to explain away (iii) since the delusions are (and some, including the DSM, say *by definition*) highly resistant to rational correction.[3] However, perhaps one can claim that the experiential evidence in favour of the delusion is so strong that this resistance to correction is not irrational since the experience trumps all possible testimony; we just don't know how strong or weird these subjects' experiences are (see Reimer 2009).

In any case, these are all contingent, descriptive issues about what these patients are actually like. One might, adopting a very different tack, turn away from what these patients are like and question whether (i-iii) rule out belief *even when they are perpetrated*. In other words, one might want to question the constitutive norms of belief being proposed by Currie and co. and claim that they are too strict. Bortolotti (2009), for example, claims that all of (i), (ii) and (iii) are too strict, and that many states that we are happy to call beliefs infringe them. She cites, in particular, the sort of biased hypothesis testing we commonly get in scientific beliefs (p.148) and also the unreviseability of racist (p.150) and religious (p.152) beliefs. She presents the anti-doxasticist with a dilemma: if we are to deny belief-status to delusions, then we are going to have to do the same for these states that, intuitively, we are happy to call beliefs. What is implicit in this argument is that denying belief-status to these non-pathological states that we are happy to call beliefs is too great a theoretical cost. Although I sympathize with the claim that Currie's norms are too strict (in particular they are plausibly not *all* constitutive norms by any means), I am unconvinced by the pivotal role accorded by Bortolotti to our actual use of the word "belief". Following, for example, Ruth Barcan Marcus (1990),

[2] De Pauw and Szulecka (1988) tell of a young man who decapitated his stepfather, taking him to be a robot, in order to look for batteries and microfilm inside his head.
[3] Indeed this feature of Capgras, has prompted a so-called "two-factor" treatment that appeals to a reasoning bias in addition to the anomalous experience (Davies, Coltheart, Langdon and Breen (2001))

we can be open to revisionism about belief.[4] As a result, if it turns out that many of the times that we say that people are believing they turn out to not really be believing, then so be it! As I said, although I don't think it is our actual use of the word "belief" that will tell us this, I do agree that Currie's norms are too strict. However, I want to show that anti-doxasticism can still be tenable even when we soften them considerably.

So, let's examine these norms, and put aside patient case studies for the moment. Which of (i), (ii) and (iii) would rule out *any* subject's being a believer of the relevant proposition? I want to show now, just how strong the anti-doxastic challenge is, by demonstrating that, even *if* we adopt very relaxed constitutive norms of belief, it is still plausible that the patient breaks them. Note that my claim is a conditional one: I am not condoning this relaxed view about the constitutive norms of belief.[5] I am adopting it for the sake of argument. Nevertheless, I do find the view attractive, and several philosophers explicitly buy into it (e.g. Schwitzgebel 2002).

The functional idiosyncracies (viz. i-iii above) divide into "upstream" and "downstream" issues. That delusions are not supported by evidence (viz. i) in their initial formation, and not responsive to review in the face of contrary evidence (viz. iii), are upstream issues (in the sense that in attempting to correct the delusion, you are providing *input* for new judgment). That they don't guide action or subsequent reasoning (viz. ii) is a downstream issue (they are *consequences* of the subject's mental state).

Suppose one were to deny that any upstream issues involve the breaking of *constitutive* norms. On such a view, upstream issues answer questions relevant to evaluative norms, about whether something is good or bad believing (e.g. questions of epistemic rationality regarding whether it is rational or justified). However, they would not be relevant to ascertaining whether it is to qualify as

[4] Such revisionism is motivated because the term in question picks out either a natural kind or an entity of fundamental theoretical importance. I think there are especially good reasons for thinking the latter with regards to belief. If belief is thought of as the fundamental informational state, the "taking of the world to be a certain way", then many cases where we ascribe beliefs to people on the basis of their claims may be cases where they don't *really* take the world to be the way their utterances suggest. One example, although there are others, involves claims (e.g. some religious claims) that are partially understood by those who utter them. We therefore might say that some Christians "believe" that "God is three and is also one". But in what sense do they really take the world to be such that *this* is the case? Such questions are, however, beyond the scope of this paper.

[5] Some convincing reasons for denying such a relaxed view come, for example, from the recent debate about the "aim of belief" (cf. Wedgwood 2002). If belief constitutively aims at something like truth or knowledge, then ignoring evidence may well preclude genuine believing.

believing at all. In other words, these upstream issues about sensitivity to evidence describe the functional role of *good* belief, not belief *tout court*. One could suggest that what interests us in this debate about whether delusions are beliefs, is whether an organism (in this case the delusional subject) takes the world to be a certain way, and if it does, what that way is and how best to characterize it. It is not strictly relevant to the question of whether the delusion is a belief (although it is clearly an interesting and important question) how it is that the organism came to have this belief (if it is indeed a belief). Indeed the idea of a mad scientist manipulating your brain so as to "implant" a certain belief is not obviously ruled out by the concept of belief.[6] This would be a prime case of irrational belief (or perhaps non-rational belief, depending on how you cash out epistemic irrationality) but nonetheless it would count as belief, for example, if it were acted upon in the right way. One might think that one doesn't need farfetched thought experiments to illustrate this point: the indoctrination of extremist cults, for example, is behaviour-driving in the worst ways (e.g. suicide bombing) and gives rise to strongly held beliefs, but the belief-formation process is far from rational.

The consequence that such a position would have for anti-doxasticism is clear. If we want to claim that delusional patients don't really believe what they assert, we need to show that the delusional patient is, all things considered, failing to act or reason in accordance with her professed beliefs. Only then should we say that she doesn't truly believe what her utterances suggest that she believes. The fact that delusions are unsupported (viz. i) and are resistant to counter-evidence (viz. iii) cannot serve to rob them of belief-status.

Now let's turn to the patients. Do they fail to act and reason in accordance with their claims? It is generally accepted that Capgras patients often fail to show concern for the replaced loved one. Patient DS (Hirstein and Ramachandran 1997) does, for example, exhibit a striking lack of concern for the welfare or whereabouts of his real parents. If, and it's not clear that this is the case, patient DS genuinely does claim "My parents have been replaced by impostors" (and not, as we shall see, merely "Those people aren't my parents"), then the answer seems to be that there is a disparity between the words and the dispositions to action. Granted, it is not always the case that *no* concern is shown for

[6] This need not be the implanting of a discrete sentence-like entity, but, for example, the "implanting" (or, better, instilling) of a particular disposition to act.

the replaced loved one. However, there is still the problem of how fleeting this concern is. Lucchelli and Spinnler (2007) give us the case of patient "Fred" and his wife "Wilma" (both, of course, fictitious names):

> On another occasion, he urged her [his wife, Wilma] to go with him to report Wilma's disappearance. *Most of the times*, however, he was quite pleased to see her as the "double" Wilma and addressed her in a very gentle way. His wife [viz. Wilma] described his manner as "courting as when we were dating" (p. 189, emphasis added).

So, although Fred occasionally acted as if his wife was missing, he was mostly perfectly happy with the double Wilma. But if the double Wilma, isn't the "real" Wilma, then where is the real Wilma? Surely these are not the actions one expects from a man who asserts that his wife, whom he loved very much, is missing.

However, one might protest that even if one puts an emphasis on downstream factors like action and dispositions to action, verbal action is action nonetheless. Asserting that p is one way of manifesting one's belief that p, one way of acting as if p were true. To this a plausible response is to say, as the saying goes, that "actions speak louder than words". Physical actions trump verbal actions because they are more of a commitment, and hence more revealing of how the subject takes things to stand in the world. If I merely assert, however sincerely, "The bridge over the ravine is safe", I don't personally risk anything if I am wrong (although I may be misinforming somebody else). If I manifest that belief by crossing the bridge, I am risking rather a lot. In a related manner, verbal action is a symbolic proxy, casting distance between the asserter and the fact asserted. I can say one thing and physically act contrary to what I have said. I can claim the bridge is safe but refuse to use it (even though, we may suppose, it is much in my interest to get across it). But I cannot physically act in two opposing ways; I cannot perform an action whose success depends on one and the same state of

affairs both obtaining and not obtaining.[7] To put it another way, when people act in ways that conflict with their professed beliefs, we take their actions to betray what they "really" believe.

Of course, extracting belief attributions in any given situation, according to the view that actions betray what is believed, requires us to have a good, clear, theory of what action is. Most relevantly, we need to distinguish action from *mere* behavior (e.g. heartbeats, reflexes, emotional responses and the like). This is a notoriously rich and tricky debate, but it is important to see that we don't need to settle this here. All we need to note is that it seems clear that the perplexing features of the delusional patient aren't behavioral: they are *agentive*. What is perplexing is not that the patient is somehow prevented from calling the police to look for his real parents or wife, it's that he doesn't *want* to, displays a lack of *concern*, and so on. In other words, borderline cases can be tricky (cf. Gendler on "alief" and belief (2008)), but this is not a borderline case.

So, to sum up the main point of this section: even if one has worries about whether the upstream functional idiosyncracies (i) and (iii) rule out belief, the anti-doxastic challenge is still on strong footing with (ii) because Capgras patients often fail to worry about the welfare or whereabouts of the replaced loved one(s) (see Davies and Coltheart 2000 for a review). For example, they won't contact the police, or make attempts to look for the "missing loved ones". Just as it is partly constitutive of your believing that something is deadly poisonous that you should refrain from eating it if you want to stay alive, so it seems to be partly constitutive of your believing that a loved one has been replaced by an impostor, that you should be concerned for the welfare or whereabouts of the loved one in question. One might say that if you are not concerned then it either can't be a "loved one" that you believe is missing (and here there are two possibilities: either you don't really believe that it is *that person* that's missing, or you can't have *loved* them very much), or you can't really *believe* that they are missing (i.e. it is some other attitude that falls short of belief, like imagining). In other words, you can locate the deviance at the level either of content or of attitude. Currie does the

[7] I can be hesitant, but here I unambiguously manifest the belief that the bridge might break. Whether epistemic modals are built into the content or the attitude is a contentious issue that I won't get into here. For a neat and plausible treatment of epistemic modals within the possible worlds framework that I favor, see Yalcin (2011).

latter and that is why he proposes that delusions are actually imaginings mistaken by the subject for beliefs.[8] In a sense, I'm proposing the latter.

With this problem firmly in mind, let us introduce the distinction between egocentric and encyclopedic doxastic states. I stress "firmly in mind" because for the next two sections there will be little mention of delusion. The reader is therefore invited to keep in mind how what follows concerning identification and belief may apply to delusional case.

2. Encyclopedic and Egocentric Doxastic States

We humans have many beliefs that we can hold in any context. This is to say that we can take the world to be a certain way, regardless of where we are and what we are experiencing. I can believe that the Battle of Hastings took place in 1066, while I am locked in a dark cupboard. These "encyclopedic" beliefs can be general such as my belief that salt dissolves in water, that lions are dangerous, or they can be singular, about individuals (or events), such as my belief that David Cameron is the current prime minister of the UK (or that the Battle of Hastings took place in 1066). I call these beliefs "encyclopedic" because they are a bit like entries in an encyclopedia. Many of these beliefs we have acquired through communicative or linguistic channels. I can change the way I take the world to be on the basis of what somebody, or some other source of information (e.g. a book or a website) tells me. Linguistically acquired belief tends to fall under the category of encyclopedic belief. I say "tends to" because, for example, verbally guiding a blind person through a maze – e.g. "To your left there is a chair" – is giving them linguistically mediated yet non-encyclopedic (what I will call egocentric) information. The information depends on the context in which the subject finds herself.

Encyclopedic beliefs that we acquire through communication come pre-packaged in a context-independent format. However, experiences of things and events we encounter can, obviously,

[8] Egan (2008) recently takes a similar approach, claiming that the delusional patient is in an in-between state of bimagining. An anti-doxasticism of a different flavour is put forward by Sass (1994), where the delusion isn't genuinely believed, but it isn't mistaken for belief either. Rather, what we have is the subject engaging in what Sass calls "a solipsistic mode of experience", which leads to "double bookkeeping". In other words, the subject doesn't locate himself in one possible world, but switches readily from one to the other, from the real to delusional realities. I put this to one side because it applies more clearly to schizophrenics than brain-damaged patients. Some of what I say in the conclusion does, however, suggest parallels with Sass's view.

be processed ("conceptualized", if you prefer) for the purposes of communication, and context-independent recall (and reasoning). In other words, we have encyclopedic beliefs – beliefs that can be held anywhere – that are *based* on a perceptual event, an event that itself is highly dependent on context.[9] I can believe the generalization that salt dissolves in water, not just because my science teacher has told me, but because I have seen it dissolve in water on numerous occasions. I can have the singular encyclopedic belief that my friend Anthony is taller than me, without anybody having to tell me this: I can see it for myself. Coming to believe these things *in a context-independent way* will require (among other things) that I have some conception of salt and of my friend Anthony, so that my beliefs come to be about these entities. After the initial perceptual encounter with these kinds or individuals of which I have a conception, once I have gleaned the information, I can in principle (e.g. memory permitting) have those beliefs anywhere, namely, while not in perceptual (or iconically/episodically recollective) contact with salt or with Anthony. This is what I mean by encyclopedic.

To ascertain whether a belief is encyclopedic, consider whether it can be believed anywhere, either without perceptual contact with the subject matter of the belief, or without recollection of a particular perceptual episode that involved the subject matter. Examples of such beliefs, as I have said, involve generalizations, such as "Salt dissolves in water". Although I may have first come to believe that on the basis of such a perceptual event, I don't need to recall or think about a particular instance of my seeing salt dissolve in water. Similarly I can believe that "My friend Anthony is taller than me" without recalling a specific instance of noticing that a man standing in front of me was taller than me. Although the content of this belief refers to me in some way (it is *my* friend who is taller than *me*), it is not, as we shall see, egocentric in the sense I will introduce now. Egocentric beliefs are not beliefs that involve reference to oneself in the content. Rather, they are beliefs that cannot be *had* independently of a perceptual or recollective episode. These perceptual and recollective episodes are implicitly egocentric in the sense that they are from a particular perspective (but the subject needn't in any way feature in the content of the belief).

[9] Although some beliefs cannot be based on perception due to the nature of the properties involved: my belief that Paris is the capital of France cannot be directly based on perception because you cannot perceive "Capitalhood".

Egocentric beliefs involve instances of taking the world to be a certain way that rely on the subject being in a certain context, experiencing (perceiving, imagining, recollecting episodically) certain things. They will often be expressible using demonstratives ("That man is [or was, in the recollective case] wearing a blue shirt").[10] Borrowing Stalnaker's (2007) terminology, I will say that these judgments haven't been "detached" from the context in which the initial judgment is made. I will now show what I mean by this.

3. From egocentric to encyclopedic belief: "detachment"

What is it for "a belief" to be "detached"? It sounds, in particular, as though there is a sense in which a belief remains the same belief once it is detached. We need to iron out an ambiguity in talking about "a belief". If we individuate a belief at the level of senses or modes of presentation then, certainly, "This man [who happens to be my friend Anthony] is taller than me" and "My friend Anthony is taller than me" are different beliefs. I can, to use the Fregean line of reasoning, rationally believe one without believing the other. For example, this could be the case if I don't know that "this man" is (or is co-referential with) "my friend Anthony". However, there is a sense in which *what is believed*, the state of affairs that is taken to be true, is the same. Both beliefs are true if and only if a certain individual is taller than I am. So, when I say "a belief is detached", this is when what is believed remains the same, but the way in which it is believed changes in a particular way, namely, from a context-dependent mode to a context-independent mode. In one sense we can think of this

[10] Aren't egocentric judgements simply indexical beliefs? Depending on what one means by "concept", there may be nothing incoherent in the notion of a concept, or conceptually mediated information, that is not encyclopaedic (indexical or perceptual concepts, or information conveyed in terms of such concepts, would be a case in point). On this view, the failure of detachability is not due to a failure of conceptualisation *altogether*, but rather due to not having conceptualised in a way that enables the information to be used in a context-independent way. When, during the perceptual experience, I think to myself "*Those* are some green eyes!" this is arguably conceptual information, but the content is egocentric insofar as its truth-conditions can only be latched onto in that perceptual context. But I think that there may also be a sense in which there can be egocentric belief, namely, the taking of the world to be a certain way, which need not rely on full-blown conceptual capacities. This is why egocentric judgements are not simply indexical beliefs, although I want to allow that the latter may be a sophisticated subset of the former. One might say that indexical beliefs involve indexical concepts, and as a result they involve conceptual content that can be merely entertained, and subsequently endorsed. With more basic egocentric judgements there may not be this separation between content and attitude: the content can be directly committal (behavior-driving) as when I dodge a lamppost.

being one belief that has changed in mode, in another sense we can think of it as a different belief. Not much hangs on this terminological ambiguity.[11]

Let us look at a normal, healthy, case of detachment. Suppose I encounter someone at a party and I judge that she has green eyes. I form the belief: "This woman has green eyes." Seeing her eyes and noticing their color has altered what I believe at that given moment, what we might call my belief-set. I've taken some information on board. Among other things, I rule out the possibility that this person's eyes are anything other than green. My judgment is correct if and only if that very person has green eyes. Now suppose I never come to know who this person is. I believed at the time that *that person* had green eyes. I might later still believe, while *physically* in a different context, that "the woman I met at the party" had green eyes. However, that isn't informationally context-independent, since, although I am no longer physically at the party, believing this relies on invoking the context of the party (perhaps in episodic memory). So it is still egocentric in the relevant sense. It is not encyclopedic in the way that my belief about David Cameron being the current prime minister of the UK is.

Now suppose my friend Tom informs me that the woman with the green eyes that I was talking to is our old friend from primary school, Alexa. My initial judgment is now detached to something that might be expressed thus: "Alexa, my old primary school friend, has green eyes." This is encyclopedic belief. Identifying this person, finding out "who they are", renders the context of the initial encounter dispensable. Not only do I no longer need to remember the context of the party, but I can forget the party altogether (i.e. be incapable of remembering the context of the party), and I could still believe that Alexa has green eyes.[12] When you know "who people are", you can have beliefs about them without needing to remember the occasion on which you got that information (this obviously goes for other, more general, beliefs too: like I said, I can't remember when I first learnt that salt dissolves in water). Knowing "who someone is", in this somewhat technical sense, clearly comes in degrees. It needn't, for example, involve knowing their name, but it must involve having

[11] To see more of this ambiguity, consider the fact that we can felicitously say of two people that they "share the belief that p", "have the same belief", "believe the same thing" etc.

[12] Notice that I can informatively communicate that fact to someone else, for example, a competent English speaker who understands to whom I am referring (perhaps he knows Alexa but hasn't noticed the color of her eyes).

enough of a conception of them (in a sense that we are about to touch upon) for you to be able to "track" them over time.[13]

What identifying Alexa actually amounts to, cognitively speaking, is that I identify the person here present with my primary school friend. I take them to be one and the same person. Some philosophers who work on theories of reference in singular thought use the metaphor of mental files (e.g. Recanati 1993) to solve various puzzles concerning singular reference. This goes back at least to Strawson (1959) who claimed that individuals (both persons and objects), which occupy a location and can change their properties, are the basic subject matter of our conceptual scheme. Mental files are also a useful tool for thinking about empirical cases of perceptual identification and tracking. In developmental psychology, there is evidence that perceptual tracking begins in very early human cognitive development (Carey and Xu 2001), plausibly prior to the development of sortal concepts. Experimental psychology has postulated low-level tracking of objects in the visual field (e.g. Pylyshyn 2007) that is independent of the perceivable properties those objects have.[14] Recently, Bullot (2009) has developed a theory of tracking (The "Empirical Tracking of Individuals" theory, or ETI for short) which postulates a mechanism for the tracking of both objects and agents. He uses a similar theoretical tool to those philosophers concerned with singular reference, which he calls a "singular file".

The general idea that I want to make use of is as follows. Each mental file corresponds to an individual, and each individual I am capable of thinking about has a file, so there are as many files as there are individuals in my mental economy. A file contains a bunch of information that I take to pertain to the individual for whom the mental file was opened. That information includes all sorts of things: biographical information, appearance properties etc. However, it is not the content of the file (or any matching of the content with something I perceive) that causes me to open or retrieve the file,

[13] Even if your conception of them is simply, "that person, whom I do know from somewhere, but I'm not sure where", it can still be said that you know "who they are" in some sense, and you will in principle be able to have information about them whose provenance is unknown (many thanks to an anonymous reviewer for pointing this out).

[14] See also Scholl 2007 for a nice review of the ties between philosophical and psychological aspects of object tracking.

but rather a non-descriptive, bare, tracking relation.[15] [16] Now, when I meet Alexa at the party, but don't know who she is, viz. haven't identified her as my old school friend, I open a new file (erroneously, as it happens, for I *have* met this individual before) for "This person here present [whom I have never met before]". She might tell me all sorts of interesting things about herself, e.g. that she has climbed Everest, and this information will go into that file (along with information about her appearance etc). All this time there is a mental file (the correct file) lying dormant (together with a large number of other files) for my old school friend, Alexa. It is "there" obviously not in the sense that if you opened my skull you would see anything resembling a file, but simply in the sense that if someone were to ask me, "Do you remember Alexa?", I would say, "Of course! We were very close at school. Her father owned a local shop and she was afraid of heights." In other words, lots of information would be easily retrieved, and retrieved together, precisely because it is taken to pertain to one and the same individual. When all is revealed and I identify this woman as my school friend Alexa, the file for "This woman here present" merges with the file for "My friend Alexa". I now take it that they are one and the same person. But this is not an explicit fact that I learn. Rather, I now "trade on" them being one and the same individual. The woman here present had a father who owned a local shop, she was afraid of heights, and went to school with me. That girl I knew went on to climb Everest (probably in the interim overcame her fear of heights) and is standing before me now.

Not much hangs on the mental files theory, or how seriously we take this notion of "files" etc., but it seems right that something roughly like this happens when we identify individuals. Identifying someone involves drawing a connection between someone currently perceived and someone one has previously encountered (it needn't be someone encountered in person, it could be someone you are familiar with from the television, e.g. "Is that Brad Pitt?!").

---

[15] A crucial feature of mental files for issues of reference, although this isn't strictly relevant here, is that the file can have largely incorrect information about the individual but still achieve reference to that individual, since it is the acquaintance relation that caused the opening of the file that fixes the referent, and not the information in the file.

[16] This is deeply reminiscent of something Pylyshyn (2007) writes: "If perceptual representations are to be grounded in the physical world then a causal link is essential at some stage in the process. The usual link that has been assumed is a semantic one – the objects that fit a particular description are the ones picked out and referred to. While this may be generally true this cannot be the whole story since it would be circular. The symbolic description must bottom out – must be grounded in objects or properties in the perceptual world. Recent evidence has suggested that the grounding is done in *objects rather than properties*".

Now, back to delusions at last. It is not for nothing that the Capgras delusion is called a delusion of mis*identification*. So, keeping in mind what identification amounts to, and the role that it plays in detaching perceptually acquired beliefs that are about individuals, let's look more closely at the Capgras delusion.

4. Consequences for the Capgras Delusion

Putting aetiological variations aside, we can agree that:

(A) The Capgras delusion has perceptual origins (it is subsequent to a perceptual encounter with a loved one that the delusion is formed: the patient doesn't, for example, simply wake up with the conviction that the loved one has been replaced).[17]

(B) That the delusional misidentification, however it may arise (e.g. whether caused or inferred on the basis of a bizarre experience) consists in an inability or reluctance to associate biographical information about the loved one with the person now present (who is *de facto* the loved one).[18] (Recall that they often acknowledge that the person they are looking at *looks exactly* like their loved one.)

The relevance that our notions of egocentric judgment and detachment have for the Capgras patient should now be clear. When I meet Alexa and fail to identify her as my old school friend, I have failed to relate this person here present to the relevant individual from my past (my old school friend Alexa), in particular to relate the two as being one and the same individual. Something analogous happens in the Capgras case. The patient fails to identify this man here present with the relevant salient individual from her past (i.e. her father). Admittedly, in the party case I fail to make the connection because Alexa has changed in appearance. The Capgras patient, on the other hand, admits that the person looks just like the person she denies that it is. This inability to provide publicly accessible reasons for belief is in large part why it is classified as a delusion.

[17] Contrast this with delusions that arise from thought disorder (e.g. in schizophrenia) where the content need have nothing to do with something perceived or even perceivable, e.g. "I am the left foot of God" (a claim made by mathematician John Nash, see, Nasar 1994).

[18] This is made nicely explicit in Lucchelli and Spinnler (2007, p.189). When Fred denies that Wilma is his wife, he cites that he "knew [her] very well as his sons' mother." Obviously, "mother of my sons" is not a perceivable property, but it is in the "Wilma file", which is failing to be correctly retrieved.

How are we to give this a personal-level characterisation in terms of beliefs and belief-sets, namely, in terms of how the person, all things considered, takes things to stand in the world? It really ought to be as easy to do this for the Capgras patient as it is for the case when I simply fail to identify Alexa at the party.[19] However, whereas at the party I act perfectly in accordance with my belief-set, the patient, as we know, fails to worry about the welfare or whereabouts of the replaced loved one. How are we to make sense of (what we might call) her agential inertia with regard to the content of her delusion?

First, let me say a word about agential inertia. By agential inertia, I mean when a subject fails to act in a way that is consistent with a belief that they have, *when that belief is taken in isolation*. As a result, agential inertia can be explained, in the sense of being rendered intelligible (as opposed to causally explicable), by other personal-level states that compete with the normal behavioral profile of a belief. To take an example at random, I may believe that somebody is a dangerous enemy, but pretend to like and trust them because I am planning to trick them in the future. This is why the likes of Quine (1960) and Davidson (1973) talk about *ceteris paribus* (other things being equal) clauses when ascribing beliefs. Note that using the *ceteris paribus* clause in cases where there are such competing states doesn't tell against the presence of the belief in question, but rather presupposes it. It presupposes that the belief is still "there" in situations where it may go un-noticed by the casual ascriber.

Now, it seems highly implausible that something of this nature can be appealed to in explaining the agential inertia of the Capgras patient's state. It is hard to think of some competing personal-level state (e.g. a motivational state) that would render the Capgras belief intelligibly inert. So instead of attempting a competing states explanation of the agential inertia, we would do well to re-examine what it is that the delusional patient might plausibly believe, in other words, to be more precise about the patient's cognitive perspective. This is where the distinction between egocentric and encyclopedic beliefs may come in handy. I'll start by making a tentative suggestion, and then I'll flesh it out. It is as follows. If we adopt a relaxed view about the constitutive norms of belief, we can hold

[19] Another potential difference is that, unlike the Capgras patient, I may have not necessarily "ruled out" that this person is Alexa. I just haven't "ruled it *in*".

something like a doxastic view with regard to the egocentric judgment that the patient might express as: "This person is not my father".[20] However, the anti-doxasticist is absolutely right that we are not in a position to accurately ascribe the encyclopedic belief that the patient might express as: "My father has been replaced by an impostor", at least on the intuitive interpretation of that sentence. It is a virtue of this position that it accounts for the following peculiarity in the patient's agential profile. Capgras patients, although they (often) fail to worry about their "missing" loved ones, do behave differently towards their (actual, present) loved ones and sometimes (as we saw, in approximately 18 % of cases) inflict violence on them. That is often, after all, why they come to the attention of the authorities. In other words, this combination of egocentric doxasticity, and encyclopedic non-doxasticity accords nicely with those cases where there is a change of general behavior towards the impostor (egocentric), coupled with a strange lack of concern for the welfare or whereabouts of the "real" loved one (encyclopedic). The immediate, constitutive, consequences of perceptually based delusional misidentification play out, but the more distant ones do not.

It remains for me to suggest why there is this combination of egocentric believing, and encyclopedic non-believing. When the patient lays eyes on her father, she fails to identify him (in mental file terms, she fails to retrieve the correct file) and judges that this man is not her father. She fails to make the link between "this man" and "my father", and takes herself, so to speak, to be in a possible world in which the man in her presence is a stranger, someone she has never met before, and hence not her father. At this point one can reasonably suppose that she is neutral about *who* this person *is* (indeed, not just neutral, but takes herself to be ignorant of that fact, since she is adamant that she has never met this person before); she is only explicitly committed to who this person *isn't*. One may even want to put this, as Bayne and Pacherie (2004) in their "endorsement theory" do, in terms of the content of her perception. In other words, what the patient's visual experience is "telling her" is something like "This person is not my father", or, perhaps even more sparsely, "I've never met this man before".

[20] Stronger constitutive norms, such as sensitivity to truth or evidence (see fn.5), may rule out a doxastic reading even in this egocentric case.

Now suppose she goes home and starts living with this man whom she judges to not be her father. She may, as is often the case, be colder towards him, make explicit claims about him not being her father. It's very troubling for those loved ones who are misidentified. And yet she shows no concern for the welfare or whereabouts of her real father. Also, as is often the case, the new conception of this man might stabilize, (in a way analogous to a scenario where, suppose, I fail to identify Alexa, but she acquires a new salience, and I think about her, that individual, under a robust but new mode of presentation, e.g. "My *new* friend Alexa") and she may even grow fond of her "new" father-figure (e.g. patient DS, "He's a nice guy, doctor, but he's not my father" (Histein and Ramachandran 1997)). (In mental file terms, another mental file, the new one, becomes stable and is filled with more and more information.) The nature of the subject's error is that she has two distinct modes of presentation (two mental files, if you will) for someone who is in actual fact one and the same individual. As far as she is concerned, there are two men, who look the same. One is her father, the man who (among other things) played a major role in her upbringing, and the other is the man she now lives with who looks like her father.

Here is how things stand. There is a fact of the matter about how this patient behaves towards her parents and we can describe it fairly unproblematically (I just have). But when we move from simply *describing* what she does (as one might describe any sequence of events) to a personal-level *explanation* of why she does it, namely, when we try to render her actions intelligible in terms of certain desires, goals and, especially, beliefs (namely, a coherent, consistent and intelligible general take on how the world is and what has happened in it) we run into very serious problems. How, we ask, can this person believe that the people she lives with are not her parents and yet fail to worry about her real parents?

Perhaps we should resist the temptation to think that belief-ascriptions neatly divide into those that are correct and incorrect. There are facts about what people believe at a given time, and what they feel and experience. These are very messy facts and are bound to be hard to catch in the net of natural language. They are even messier when the people in question have serious brain damage. We should instead think that some ascriptions will be more accurate in capturing the subject's cognitive perspective than others, and will therefore be better at rendering her actions intelligible. I

would suggest that egocentric ones will generally be more "accurate" (in this sense) than encyclopedic ones. Compare the following three belief ascriptions (presented as how the subject would verbally express them).

- "This man, whom I am looking at now, is not my father". This is egocentric, and seems accurate. The patient certainly does not act towards him as if he *were* her father (cf. violence, coldness, no mention of shared past experiences etc.).
- "The man I live with is not my father". This is an (erroneously detached) encyclopedic belief. She can believe this anywhere. It seems fairly accurate as an ascription. We certainly wouldn't want to say that she believes that she *is* living with her father.
- "My real father has been kidnapped, and replaced by an impostor". This, which is the sort of thing that is commonly ascribed to the Capgras patient, ("The Capgras patient believes that her loved ones have been replaced by impostors") does not seem to capture the patient's overall mental state at all. If she *did* believe this, she would have to show concern for the welfare of her real parents, which we know she fails to do.

5. Conclusion

The main lesson is that we need to be very careful about what we ascribe to the patient in terms of beliefs and believing. We can see this, not so much as criticizing either of the doxasticist or anti-doxasticist positions, but rather as changing the terms of the debate. The debate has traditionally asked: "Are delusions beliefs?" By asking this there have been a number of implicit assumptions, the clearest of these being that there is a discrete, belief-like state that has the same content as the patient's delusional assertions (which is also taken to match up with the standard clinical characterization of the delusion: "that a loved one has been replaced by an identical-looking impostor").[21] The challenge is then to ascertain whether the state in question has the right functional role. A consequence of this is that the *content* of the delusional state is not questioned (it is treated a bit like a sentence in the head); rather it is the attitude that the subject bears towards the content that is

[21] In fact transcripts indicate that there is lots of use of demonstrative reference to the perceived impostor: "*That man* is not my father".

up for dispute. This leads some, like Currie and his collaborators, to claim that delusions are imaginings that are mistaken for beliefs by the subject (that patient imagines that p and so asserts that p, but fails to genuinely believe that p). Others, like Egan (2008) – who himself openly admits that he is *presupposing* both functionalism, and the representational theory of mind – blur the traditional boundaries between attitudes and claim that delusions are in-between states (bimagining; somewhere between belief and imagining).[22] One option is to accept this starting point and the conclusions that follow. But an alternative is to view this as a *reductio*: these conclusions are unsatisfactory to the point of being an indication that we aren't asking the right questions or approaching the issue in the right way. If instead we ask: "How does the patient take things to stand in the world?", or in Lewis/Stalnaker terminology, "What possible world does that delusional subject take herself to be inhabiting?" we get to the heart of the problem, without implicitly subscribing to any presuppositions about what believing is and what underpins it. The real problem with delusional patients is that we find it hard to grasp how they take the world to be. The question of whether a fixed representational state, "the delusion", is a belief or not is secondary, and comes with a great deal of theoretical baggage.

References

Bayne, T. and Pacherie E. (2004). "Bottom up or top down?" *Philosophy, Psychiatry, & Psychology*, 11 (1): 1–11.

Bayne, T. and Pacherie, E. (2005). "In defence of the doxastic conception of delusion," *Mind & Language*, 20 (2): 163–188

Bortolotti, L. (2009). *Delusions and Other Irrational Beliefs*, Oxford: Oxford University Press.

[22] Here is Egan 2008 (I quote this in full because it is a wonderfully clear and explicit expression of the implicit orthodoxy in this debate): "In what follows, I'll be concerned with characterizing (very partially) the attitude that delusional subjects bear to the contents of their delusions. I'm going to assume that the right way to go about this is to characterize the roles that particular, token mental representations play in the subject's cognitive economy. So I'll be assuming that some sort of minimal representational theory of mind is correct – that there is a medium of mental representation, and there are discrete representational items in the head. These representational items are operated on in various ways, and accessed by various systems, in order to regulate both our behavior and the maintenance of other representational items. Believing, desiring, imagining, and the bearing of propositional attitudes in general is a matter of having a representational item with the right kind of content, which plays the right kind of role in one's cognitive economy." (p. 262)

Bullot, N. (2009). "Toward a Theory of the Empirical Tracking of Individuals: Cognitive Flexibility and the Functions of Attention in Integrated Tracking". *Philosophical Psychology* 22 (3):353-387.

Carey, S. and Xu, F. "Infants' Knowledge of Objects: Beyond Object Files and Object Tracking", *Cognition* 80 (2001) 179-213

Currie, G. (2000). "Imagination, delusion and hallucinations," in M. Coltheart and M. Davies (eds.) *Pathologies of Belief*, Oxford: Blackwell, 167–182.

Currie, G. and Jureidini, J. (2001). "Delusions, rationality, empathy," *Philosophy, Psychiatry and Psychology*, 8 (2–3): 159–162.

Davidson, D. (1973), "The Material Mind", Reprinted in Davidson (1980), *Essays on Actions and Events*, Oxford: Clarendon Press

Davies, M., Coltheart, M., Langdon, R. and Breen, N. (2001). "Monothematic delusions: Towards a two- factor account," *Philosophy, Psychiatry and Psychology*, 8(2/3): 133–158.

de Pauw, K.W., Szulecka, T.K. (1988) "Dangerous delusions. Violence and the misidentification syndromes." *British Journal of Psychiatry* 152: 91–6

Egan, A. (2008). "Imagination, Delusion and Self-Deception" in T. Bayne and J. Fernàndez (eds.), *Delusion and Self deception: Affective and Motivational Influences on Belief Formation*, Hove: Psychology Press

Ellis, H. D., Young, A.W. (1990). "Accounting for delusional misidentifications". British Journal of Psychiatry, 157: 239-248.

Hirstein, W. & Ramachandran, V. S. (1997). "Capgras syndrome: a novel probe for understanding the neural representation of the identity and familiarity of persons." *Proceedings of the Royal Society B: Biological Sciences*, *264*, 437-444.

Lucchelli, F. and Spinnler, H. (2007). "The case of lost Wilma: a clinical report of Capgras delusion," *Neurological Science*, 28(4): 188–195.

Marcus, Ruth B. (1990), "Some revisionary proposals about belief and believing", *Philosophy and Phenomenological Research*, 50, 132–153.

Maher, B.A. (1974). "Delusional thinking and perceptual disorder," *Journal of Individual Psychology*, 30: 98–113.

Nasar, S. (1994). *A Beautiful Mind: A Biography of John Forbes Nash, Jr., Winner of the Nobel Prize in Economics*. Simon & Schuster.

Quine, W.V.O., (1960), *Word and Object*, Cambridge, Massachusetts: MIT Press.

Recanati, F. (1993). *Direct Reference*. Cambridge University Press.

Sass, L. (1994). *The Paradoxes of Delusion: Wittgenstein, Schreber, and the Schizophrenic*, *Mind*, Ithaca: Cornell University Press.

Scholl, B. (2007). Object Persistence in Philosophy and Psychology. *Mind and Language* 22 (5):563–591.

Schwitzgebel, E. (2002) “A phenomenal, dispositional account of belief”, *Nous*, 36, 249–275.

Stalnaker, R. (2007). *Our Knowledge of the Internal World*. Oxford University Press.

Strawson, P., (1959). *Individuals*, London: Methuen.

Wedgwood, R. (2002), “The Aim of Belief”, *Philosophical Perspectives* 16, 267–297.

Yalcin, S. (2011). “Nonfactualism About Epistemic Modality”. In A. Egan & B. Weatherson (eds.), *Epistemic Modality*. Oxford University Press.